# Rearrangement of the Extracellular Domain/Extracellular Loop 1 Interface Is Critical for Thyrotropin Receptor Activation*[S]

Joerg Schaarschmidt[‡1], Marcus B. M. Nagel[‡§1], Sandra Huth[‡], Holger Jaeschke[‡], Rocco Moretti[¶], Vera Hintze[||], Martin von Bergen[§**‡‡], Stefan Kalkhof[§§§], Jens Meiler[¶], and Ralf Paschke[||||1]

From the ‡Department of Internal Medicine, University of Leipzig, 04103 Leipzig, Germany, §Department of Proteomics, Helmholtz-Centre for Environmental Research, 04318 Leipzig, Germany, ¶Department of Chemistry and Center for Structural Biology, Vanderbilt University, Nashville, Tennessee 37235, ||Institute of Materials Science, Max Bergmann Center of Biomaterials, TU Dresden, 01069 Dresden, Germany, **Faculty of Biosciences, Pharmacy and Psychology, Institute of Biochemistry, University of Leipzig, 04103 Leipzig, Germany, ‡‡Aalborg University, Department of Chemistry and Bioscience, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark, §§Department of Bioanalytics, University of Applied Sciences and Arts of Coburg, 96450 Coburg, Germany, and ||||Division of Endocrinology and Metabolism and Arnie Charbonneau Cancer Institute University of Calgary, Calgary, Alberta T2N 1N4, Canada

The thyroid stimulating hormone receptor (TSHR) is a G protein-coupled receptor (GPCR) with a characteristic large extracellular domain (ECD). TSHR activation is initiated by binding of the hormone ligand TSH to the ECD. How the extracellular binding event triggers the conformational changes in the transmembrane domain (TMD) necessary for intracellular G protein activation is poorly understood. To gain insight in this process, the knowledge on the relative positioning of ECD and TMD and the conformation of the linker region at the interface of ECD and TMD are of particular importance. To generate a structural model for the TSHR we applied an integrated structural biology approach combining computational techniques with experimental data. Chemical cross-linking followed by mass spectrometry yielded 17 unique distance restraints within the ECD of the TSHR, its ligand TSH, and the hormone-receptor complex. These structural restraints generally confirm the expected binding mode of TSH to the ECD as well as the general fold of the domains and were used to guide homology modeling of the ECD. Functional characterization of TSHR mutants confirms the previously suggested close proximity of Ser-281 and Ile-486 within the TSHR. Rigidifying this contact permanently with a disulfide bridge disrupts ligand-induced receptor activation and indicates that rearrangement of the ECD/extracellular loop 1 (ECL1) interface is a critical step in receptor activation. The experimentally verified contact of Ser-281 (ECD) and Ile-486 (TMD) was subsequently utilized in docking homology models of the ECD and the TMD to create a full-length model of a glycoprotein hormone receptor.

Glycoprotein hormones (GPHs)[3] normally regulate crucial processes in metabolism and reproduction by activating GPHRs. This is especially true for TSHR, which can cause several clinically relevant conditions like hypo- and hyperthyroidism when it malfunctions. Yet the mechanism of how extracellular ligand binding induces the structural changes required for intracellular G protein activation is unknown. We pursued an integrated structural biology approach using modeling guided by experimental data to generate experimentally supported full-length TSHR models. It is expected that some insights gleaned from a TSHR model can be generalized to other GPHRs. These models in turn create testable hypotheses on the mechanism of GPHR activation and can promote drug development to treat GPHR-associated diseases.

GPHs bind to the ECD of their respective receptors (Fig. 1A) and consequently initiate activation, which is presumably propagated by induction of conformational changes within the ECD's hinge region (HR) (1–4). Interestingly, GPHRs still possess a binding site within the TMD not associated with physiological receptor activation but accessible to low molecular weight agonists and allosteric modulators (5–7). Another important aspect of GPHR function and physiology is post-translational modification, including disulfide bond formation, glycosylation, tyrosine sulfation, and proteolytic cleavage with the latter only occurring during maturation of the TSHR (for review, see Kursawe et al. (8)). However, there is no evident physiological requirement for proteolytic excision of the ~50 amino acid C-peptide, with a deletion variant showing similar characteristics to the wild type (WT) receptor (9). In contrast, glycosylation and sulfation are obligatory, with the latter being

---

[S] This article contains supplemental models.
[1] Both authors contributed equally to this work.
[2] To whom correspondence should be addressed: Richmond Road Diagnosis and Treatment Centre (RRDTC), University of Calgary, 1820 Richmond Rd. SW, Calgary, Alberta T2T5C7, Canada. Tel.: 403-955-8969; Fax: 403-955-8248; E-mail: ralf.paschke@ucalgary.ca.

[3] The abbreviations used are: GPH, glycoprotein hormone; TSHR, thyroid stimulating hormone (TSH) receptor; bTSH, bovine TSH; GPCR, G protein-coupled receptor; ECD, extracellular domain; HR, hinge region; FSHR, follicle stimulating hormone receptor; XL, cross-linker; LRA, linear regression analysis; DST, disuccinimidyl tartrate; BS3, bis(sulfosuccinimidyl) suberate-D0/D4; sulfo-EGS, ethylene glycol bis(sulfosuccinimidylsuccinate); BS(PEG)5, PEGylated bis(sulfosuccinimidyl)suberate; ESI, electrospray ionization; CAM, constitutively activating mutations; TMD, transmembrane domain; GPHR, glycoprotein hormone receptor.
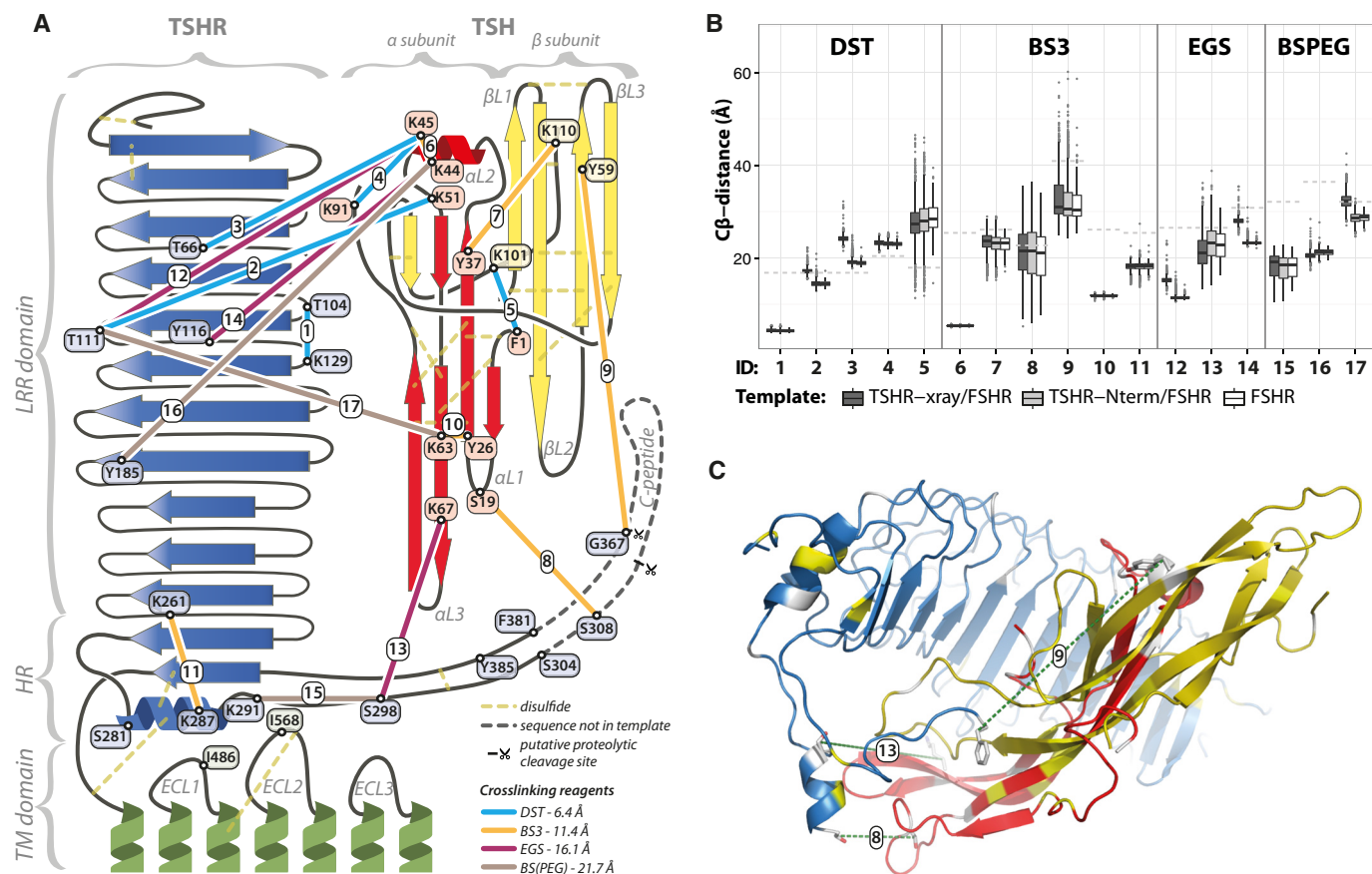
FIGURE 1. **Schematic representation and identified cross-links of the TSHR·TSH complex.** *A*, schematic representation of the TSHR·TSH complex including disulfides, cross-linked residues identified by mass spectrometry and significant residues of the TSHR, including residues with reported constitutively activating mutations (Ser-281, Ile-486, Ile-568), the sulfation site (Tyr-385), and boundaries of the model within the HR (Phe-381, Ser-304). The respective spacer-length of the cross-linking reagents is specified in the figure legend. *B*, boxplot of Cβ distance distribution between residues connected by chemical cross-linking within the homology models of the TSHR-ECD·TSH complex. The employed cross-link-specific cutoff distance (Table 1) is indicated by a *dashed line in gray*. Cross-links 8 and 9 include one residue located in the part of the HR not included within the models. For these, the distance to the closest residue included in the models is reported, and the missing residues are considered in the cutoff distance. *C*, cross-links (*green lines*) between the hinge region of the TSHR (*blue*) and the hormone (*red*, α-chain; *yellow*, β-chain) suggest that the HR, including the part not resolved in the FSHR-ECD/FSH template and, therefore, not included in the homology models, is oriented toward the hormone and most likely also contributes to ligand binding.

an indispensable feature of specific hormone binding (8, 10). The structure of the ECD of the follicle stimulating hormone receptor (FSHR, a member of the GPHR subfamily) in complex with FSH (11) showed that the ECD forms a continuous hand-shaped structure. In this the C-terminal HR does not form a separate structural entity as previously anticipated but, rather, comprises the last two β-sheets of the LRR-fold, an α-helix as well as the "thumb" region including the sulfation located at the interface to the hormone. Despite these invaluable insights on ligand binding and specificity, many details about GPHR activation are still elusive. These include the potential role of the HR residues with unresolved electron density, the significance of receptor oligomerization, and negative cooperativity in hormone binding (4). A major obstacle in understanding GPHR activation has been the lack of an atomic detail model, particularly one that defines the relative orientation of ECD and TMD, identifies interacting residues at the interface, and illustrates the structural changes upon ligand binding within the HR and the interface.

In pursuit of a full-length structural GPHR model we implement an integrated computational/experimental approach. Chemical cross-linkers (XL) of a defined maximal length react intra- or intermolecularly with two functional groups on the protein surface. After enzymatic digestion, the resulting fragments are identified by mass spectrometry (MS). Based on the spacer lengths, an approximate upper boundary for the distance is derived and employed as a restraint for the structural models (12). This approach is limited to the soluble ECD because of difficulties purifying a functional, full-length TSHR, even in the very low quantities needed for cross-linking experiments (11, 13). Therefore, we additionally use double-mutant cycle analysis and disulfide cross-linking to assess the direct contact between amino acids at the ECD/TMD interface (14, 15). Even though resulting structural restraints are sparse, they are sufficient to build structural models for the full-length TSHR with the Rosetta software suite (16). These models provide insights to TSHR activation. Specifically, we predict the relative orientation of ECD and TMD, potentially important contact points at the ECD/TMD interface, and the conformational changes necessary for receptor activation. The high sequence conservation of the investigated region within the GPHR subfamily as well as studies on chimeric receptors (17, 18) suggest a shared activation mechanism with receptor-specific interactions. The reported approach can, therefore, be

expanded to the remaining GPHRs and provide new insights into similarities as well as receptor-specific features of GPHR activation.

## Experimental Procedures

*Purification of the Soluble TSHR-ECD*—A soluble TSHR-ECD with a 10-histidine tag and a glycosylphosphatidylinositol (GPI) anchor (TSHR_ECD10HisGPI) was expressed and purified as previously described (13). Briefly, the gene was stably transfected into CHO Flp-In™ cells (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. Purification was performed by liquid chromatography at 4 °C with a nickel-Sepharose high performance affinity column (HisTrap HP 5 ml; GE Healthcare). After column equilibration the sample was applied (flow rate: 0.5 ml/min), and the collected fractions were tested for presence and purity of the soluble ECD by SDS-PAGE followed by Coomassie staining or Western blotting with anti-TSH receptor antibody (A9, Abcam®). Fractions containing the ECD in sufficient quantity and purity (>70%) were combined, concentrated, and buffer-exchanged with PBS with a centrifugal concentrator (Corning® Spin-X® UF 20 ml, molecular weight cutoff 10).

*Cell Culture, Transient Expression, and Characterization of Wild Type and Mutant Full-length TSHR*—Mutations were introduced into the human TSHR gene and tagged with an N-terminal hemagglutinin tag in a pcDNA3.1(-)/hygromycin vector via site-directed mutagenesis as described previously (6). COS-7 cells were then transiently transfected with the WT and mutated vectors using the GeneJammer transfection reagent (Stratagene, Amsterdam, The Netherlands). Functionality of expressed TSHR variants was evaluated as described previously (19, 20) by determining cell surface expression, specific binding of bovine TSH (bTSH, National Hormone and Pituitary Program of the NIDDK, National Institutes of Health), basal- and bTSH (30 milliunits)-induced cAMP accumulation, and linear regression analysis (LRA) of basal cAMP accumulation *versus* cell surface expression.

The $G_{q/11}$ activation was determined in $HEK_{GT}$ cells by cotransfection of the vectors with a reporter vector harboring the firefly luciferase gene under the control of the nuclear factor of activated T-cells (NFAT) transcription factor (pNFAT-Luc, Agilent Technologies, Santa Clara, CA). 48 h after transfection cells were stimulated for 4 h with bTSH (30 milliunits) and lysed with Luciferase Cell Culture Lysis Reagent (Promega, Madison, WI). Luciferase activity was determined as described previously by Hampf and Gossen (21).

*MS Analysis of Cross-linked Soluble TSHR-ECD*—Chemical cross-linking of the soluble TSHR-ECD with bTSH was performed as previously described (12, 22) with disuccinimidyl tartrate (DST), bis(sulfosuccinimidyl) suberate-D0/D4 (BS3), ethylene glycol bis(sulfosuccinimidylsuccinate) (sulfo-EGS), and PEGylated bis(sulfosuccinimidyl)suberate (BS(PEG)5) obtained from Thermo Fisher Scientific Inc. (Rockford, IL) as well as BS3-D0/D12 obtained from Creative Molecules Inc. (Toronto, OLN, Canada). All cross-linking reactions were conducted in 1× PBS buffer, pH 7.2, at a protein concentration of 3.3 or 2.5 $\mu$M. The molar ratio of protein to cross-linker was 1:100 or 1:200, respectively, and the reaction was quenched after 60 and 120 min with ammonium bicarbonate buffer equimolar to the cross-linking reagent. The cross-linking reagents BS3 and sulfo-EGS were resolved in 1× PBS at a final concentration of 0.1 mol/liter. Hydrophobic cross-linking reagents BS(PEG)$_5$ and DST were freshly prepared in dry DMSO to a final concentration of 0.25 mol/liter and 0.1 mol/liter, respectively.

The cross-linked proteins were deglycosylated with 250 units of peptide *N*-glycosidase F according to the manufacturer's instructions. The samples were subsequently separated by gradient SDS-PAGE (4–12%). Bands at the positions corresponding to the molecular weights of TSHR-ECD, TSH, and the complex were excised, and samples were reduced, alkylated, and digested in-gel using trypsin.

The resulting peptide mixtures were analyzed by nano-HPLC/nano-ESI-MS/MS using either an Orbitrap Fusion Tribrid or LTQ Orbitrap XL ETD mass spectrometer. Identification was performed using the software StavroX (23), allowing a mass deviation of 15 ppm (LTQ Orbitrap) or 10 ppm (Orbitrap Fusion) for MS precursors and 0.5 Da (collision-induced dissociation, LTQ Orbitrap XL ETD) or 0.1 Da (higher energy collision dissociation, Orbitrap Fusion Tribrid) for fragment ions.

*Surface Plasmon Resonance*—Surface plasmon resonance was performed on a T100 (Biacore, Uppsala, UC, Sweden). Recombinant TSHR-ECD was amine-coupled on a CM3-Chip following standard procedures. The final protein loaded amounted to 210 relative units. Experiments were conducted for eight different ligand concentrations (1500, 500, 166.67, 55.56, 18.52, 6.17, 2.06, and 0 nM) at a flow rate of 30 $\mu$l/min and 25 °C. Contact time of the ligand was 300 s followed by 800-s dissociation time. The regeneration was performed using 2.5 M NaCl in HBS-EP (GE Healthcare) for 30 s followed by 200 s for stabilization. Data analysis was performed using Sigma Plot 12.0 (Systat Software Inc, Bangalore, Karnataka, India) and Biacore T100 evaluation Software 2.03.

*Nano-HPLC/NanoESI-LTQ Orbitrap XL ETD MS*—Samples were prepared in 0.1% formic acid, injected in a NanoAcquidity UPLC, trapped, and desalted for 10 min on a $C_{18}$ trapping column (nanoACQUITY symmetry trapping column, Waters) with a constant flow of 15 $\mu$l/min and 2% acetonitrile. After 8 min the peptides were eluted and separated on a $C_{18}$ reverse phased column (ACQUITY UPLC Peptide BEH C18 nanoAC-QUITY, Waters) using a linear acetonitrile gradient (8–45%) over 85 min or 140 min (Waters Corp., Milford, MA) at a flow rate of 300 nl/min. The HPLC system was coupled online to a mass spectrometer via a chip-based nanoESI source (TriVersa NanoMate, Advion, Ithaca, NY). The spray voltage was set to 1.6–1.8 kV, and the capillary was heated to 250 °C. MS/MS-scans were triggered automatically after each full scan (*m/z* range of 400–2000, resolution of 60,000, 1 microscan, and 5 × $10^5$ ions accumulated) for the 6 or 10 highest abundant precursor ions, exceeding an intensity of $10^3$ and a charge state of ≥2. The employed lock mass for online recalibration was 445.1200 *m/z*. Furthermore, the instrument was set to exclude ions from a dynamic exclusion list (500 entries) with a maximal retention period of 60 s and a relative mass window of ±3 Da for MS/MS scans. Fragmentation of selected precursor ions ±4 Da was

caused by collision-induced dissociation with ramped normalized collision energy of 37 ± 15 (three steps). Activation Energy (Q) was set to 0.250 with an activation time of 30 ms. The automatic gain control target was set to 8000 ions, and the fragment analysis took place in the ion trap.

*Nano-HPLC/NanoESI-Orbitrap Fusion Tribrid MS*—Samples were prepared in 0.1% formic acid, injected in an UltiMate 300 HPLC, trapped, and desalted for 8 min on a $C_{18}$ column (Acclaim PepMap100) with a constant flow of 5 $\mu$l/min and 2% acetonitrile. Afterward peptides were eluted and separated on a $C_{18}$ separation column (Acclaim PepMap RSLC column) using a linear acetonitrile gradient (8–45%) over 80 or 130 min (Dionex Corp., Sunnyvale, CA) at a flow rate of 300 nl/min. The HPLC system was coupled online to a mass spectrometer via a chip-based nanoESI source (TriVersa NanoMate, Advion). The spray voltage was set to 1.7–1.8 kV, and the capillary was heated to 275 °C. MS/MS-scans were triggered automatically after each full scan (*m/z* range of 350–2000, a resolution of 60,000, 1 microscan, and $5 \times 10^5$ ions accumulated) using a top speed decision tree (5-s cycle time) setting the highest priority for the highest charge state followed by the highest abundance. Precursor ion intensity was required to exceed $2 \times 10^3$, and the charge state was restricted to a range of 2–7 *m/z*. The employed lock mass was 445.1200 *m/z*. The instrument was set to exclude ions from a dynamic exclusion list with a maximal retention period of 15 s and a relative mass window of ±20 ppm for MS/MS scans. Fragmentation of selected precursor ions ±4 Da was caused by higher energy collision dissociation with stepped normalized collision energy of 35 ± 10. The automatic gain control target was set to 5,000 ions. Fragment ions were detected in the Orbitrap at a resolution of 15,000.

*Molecular Modeling of the Full-length TSHR*—The homology model of the TSHR-ECD in complex with bovine TSH was generated using Rosetta 3 (16, 24). Briefly, homology modeling was based on the structure of the FSHR-ECD in complex with FSH (Ref. 11; PDB ID 4ay9). In addition, sections of the LRR domain were replaced by the coordinates of the TSHR-LRR domain (Ref. 25; PDB ID 2xwt) after superimposing the residues at the junctions (cut after Leu-57 or Ser-234 of the TSHR-LRR domain). The protein sequences of the TSHR-ECD and TSH were subsequently aligned to the structural coordinates of the template structures. For each template a set of 2000 models (150 for the FSHR template) was built, reconstructing backbone coordinates in gapped regions of the alignment using the cyclic coordinate descent (CCD) protocol followed by a relaxation of the structures after side-chain coordinates were added from a rotamer library. The structures were clustered using Calibur (26). In addition C$\beta$ distances for each model were determined with Rosetta's contactMap protocol (20).

Homology models of the TSHR-TMD were generated with the RosettaCM protocol as described by Song *et al.* (27). Homology modeling was performed for 20 templates of class A GPCRs considered to be in an inactive conformation (PDB codes 4n6h, 2rh1, 3uon, 4ej4, 4eiy, 1u19, 3rze, 4bvn, 4dkl, 2z73, 4u15, 4djh, 4ea3, 3v2y, 4mbs, 4ib4, 3pbl, 4ntj, and 3odu) and seven templates considered to be in an active conformation (4lde, 4mqs, 2ydv, 4j4q, 2y00, 3ayn, 4iar). A sequence and structure-based alignment of the templates was performed with the

Molecular Operating Environment (MOE, 2012.10; Chemical Computing Group Inc., Montreal, QC, Canada) with manual adjustment of the alignment removing gaps within the core TM regions. For each template set 5000 models were generated, and the resulting models were clustered with Calibur both as separate sets and combined.

All homology models were evaluated based on energy and cluster size. The best scoring model from each of the 5 largest ECD model clusters was subsequently docked to the three best scoring models from each of the 10 largest TM domain clusters. Before the docking run the C-terminal loop segment of the ECD, which most likely adopts an unrepresentative conformation in the models due to the missing TMD, and the ligand bTSH were removed. In the initial placement of the two partners the ECD was placed arbitrarily in an upright position above the interface with the TMD. The initial perturbation included a random spin between 0 and 360° around, and a random tilt between 0 and 90° along the sliding axis (roughly parallel to the membrane normal). For this purpose the tilt option was implemented and incorporated into the Rosetta Software suite, allowing a random tilt within a predefined limit during the initial perturbation step of the docking protocol (28). During docking a cross interface disulfide between Cys-284 and Cys-408 was enforced. Furthermore, the low resolution step of the docking protocol was repeated until the C$\beta$ distance between Cys-284 and Cys-408 was <15 Å. For each ECD/TMD combination 1000 models were built. The Ser-281/Ile-486 C$\beta$ distance for each model was determined with Rosetta's contactMap protocol and the interface energies with the InterfaceAnalyzerMover (29). For remodeling of the linker region (Lys-401–Ile-411) two sets were selected; (i) all models of the best 100 by dG separated with a Ser-281–Ile-486 C$\beta$ distance of <15 Å (57 models) and (ii) all models with dG separated < −6 and a Ser-281–Ile-486 C$\beta$ distance of <10 Å (41 models). For each of these, 25 loop models were generated with a subsequent relaxation step. The resulting models were again clustered with Calibur. The best scoring models of the 10 largest clusters have been deposited at the model archive (Model Archive database; 10.5452/ma-aptif). Contact maps were generated for each cluster as well as for all models with a C$\beta$ distance cutoff of 8 Å. For the best scoring structure of each cluster the position of the hormone and the position and environment of the sulfated tyrosine residue recapitulated those of the initial homology model of the extracellular domain. The junctions were remodeled (25 decoys) followed by relaxation of the entire structure. Visualization and image generation was done using the PyMOL Molecular Graphics System (Version 1.5.0.4 Schrödinger, LLC).

## Results and Discussion

### Strategy for TSHR Structure Prediction Based on Chemical Cross-linking and Mutation Data

Structural modeling of the full-length TSHR was based on structural templates resolved by x-ray crystallography of the GPHR-ECD (11, 25) as well as the TMD of class A GPCRs (Fig. 2). Experimental data from chemical cross-linking of the soluble TSHR-ECD with bTSH was incorporated to guide and eval-
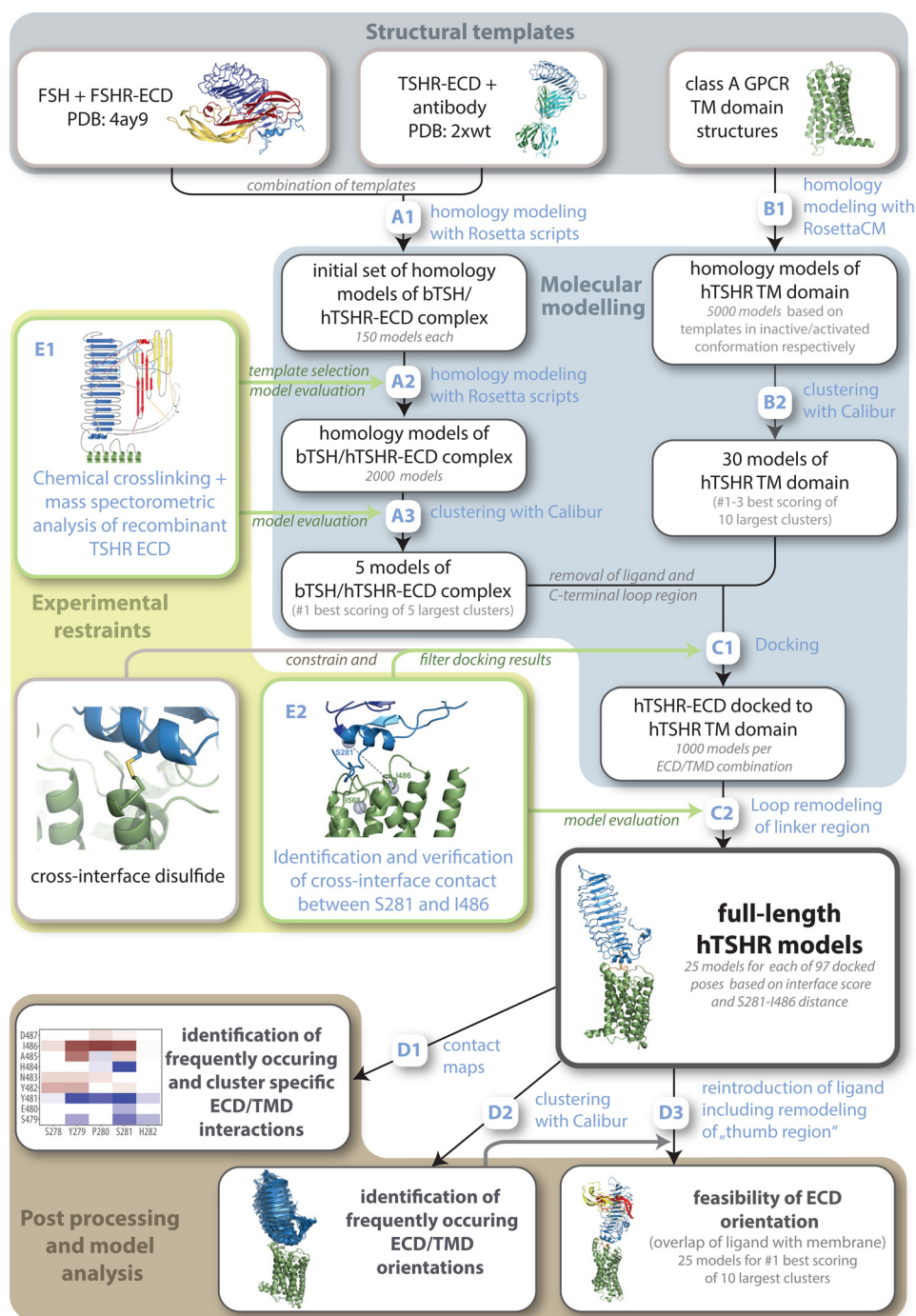
FIGURE 2. **Strategy for generating full-length GPHR models.** An integrated structural biology approach combining computational techniques (*A–D*) with experimental data (*E1* and *E2*). Homology models were constructed using Rosetta 3 for the TSHR-ECD (*A1* and *A2*) and the multitemplate approach of RosettaCM for the TSHR-TMD (*B1*). Chemical cross-linking of the soluble TSHR-ECD yielded 17 cross-links that were used to guide template selection and evaluate the models of the TSHR-ECD (*E1*). The model sets were further analyzed by clustering analysis using Calibur (*A3* and *B2*). Models were selected based on energy and cluster size. The combination of 30 TSHR-TMD models with 5 TSHR-ECD models by docking yielded 150,000 docked models (*C1*). During docking a cross-interface disulfide between Cys-284 and Cys-408 was enforced. From the docked poses ~100 models were selected based on interface score and agreement with the experimentally verified contact of Ser-281 with Ile-486 (*E2*) for reconstruction of the linker region (Lys-401–Ile-411, *C2*). The model set of the full-length TSHR was further analyzed by contact maps (*D1*) and clustering (*D2*). Feasibility of the full-length models was verified by reintroduction of the ligand and remodeling of the thumb region (*D2*).

uate the homology modeling of the TSHR-ECD·TSH complex. A number of class A GPCR experimental structures were incorporated into homology modeling of the TSHR-TMD by utilizing the multiple template approach of RosettaCM (27). The models of the ECD and the TMD were combined by docking with subsequent remodeling of the linker region. In this step

structural flexibility of the interfaces was incorporated by combining various homology models of the ECD and TMD during docking. The putative contact of the ECD residue Ser-281 with the TMD was identified and verified by double mutant cycle analysis. This contact was used to guide and evaluate placement of the ECD in relation to the TMD during docking along with

the cross-interface disulfide (11, 30). To gain information on frequently occurring ECD/TMD orientations and specific interface contacts, the final ensemble of models was analyzed by clustering and contact maps. Plausibility of the most frequent ECD/TMD orientations was verified by reintroduction of the ligand into the models.

### Mass Spectrometry Analysis Confirms Glycosylation, Sulfation, and Proteolytic Cleavage of the Extracellular Domain

*Glycosylation*—Mass spectrometric analysis of the soluble TSHR-ECD after tryptic digestion identified fragments covering 80% of the protein sequence of the utilized construct. The analysis of glycosylation sites revealed complete glycosylation of three of the six putative sites within the ECD, at Asn-77/-113/-177 and a partial glycosylation of Asn-302 (31). Due to the absence of detected proteolytic peptides covering the remaining two sites (Asn-99/-198), glycosylation of these sites could not be determined by MS.

*Sulfation*—Sulfation of the TSHR-ECD was identified at position Tyr-385, as suggested by Costagliola *et al.* (10) as well as at position Tyr-387. Sulfation was typically identified at a single site or at both sites simultaneously; the peptide representing the non-sulfated form of TSH was rarely observed. However, the mutagenesis data clearly show that the functional importance of tyrosine sulfation is exclusively attributed to Tyr-385, with no functional compensation by Tyr-387. However, whereas sulfation of Tyr-385 and Tyr-387 was determined in a truncated ECD, functional data were gathered from the full-length receptor. Given that secondary structure supposedly has a major influence on sulfation (32) and with the structural influence of the TMD on the HR (17), there might be a discrepancy between sulfation of the truncated and the full-length TSHR, with sulfation of Tyr-387 occurring only in the truncated receptor.

*Proteolytic Cleavage*—Wadsworth *et al.* (33) suggested that residues Ala-317–Phe-366 are posttranslationally removed with no apparent effect on TSHR function. Analysis of proteolytic cleavage of the TSHR by MS confirmed C-terminal cleavage between position Phe-366 and Gly-367 by detection of a proteolytic peptide (Gly-367–Lys-371). Yet no peptide confirming the N-terminal cleavage site between Asn-316 and Ala-317 was detected. Recent studies suggest that excision occurs by successive removal of small fragments, resulting in ragged boundaries (34–36).

### Homology Models of the TSHR-ECD·TSH Complex Consistent with Chemical Cross-linking Data

The TSHR-ECD·TSH complex was studied by chemical cross-linking and MS using four different amino-reactive cross-linking reagents with differing spacer lengths. Before cross-linking the tight binding of bovine TSH to the TSHR-ECD was verified by surface plasmon resonance spectroscopy. Steady-state analysis indicates a two-site binding model as previously reported (13). Seventeen unique distance restraints could be determined within the TSHR-ECD·TSH complex (Fig. 1 and Table 1). These included nine receptor-hormone cross-links, two cross-links between the subunits of the hormone, and three within the receptor and the α-chain of the hormone, respec-

**TABLE 1**

**Cross-linked peptides of the TSHR-ECD·TSH complex identified by mass spectrometry and resulting estimated cutoff distances for structural modeling**

The suffixes after residue numbers indicate the component of the complex to which the residue belongs (a, α-chain of TSH; b, β-chain of TSH; r, receptor). The covalently cross-linked residue of the peptide is highlighted in bold. If the cross-linking reaction occurred within the protein N-terminal amino group, a bracket is added. In cases where the data are ambiguous, the less likely cross-linking sites are underlined. Modified residues are annotated in the sequence (D*, deaminated asparagine; m, oxidized methionine; B, cysteine acetamide). u, unified atomic mass.

| ID | 1st residue | 2nd residue | 1st Sequence | 2nd Sequence | Cross-linker | MS/MS | Δm (ppm) | MH+ experimental (u) | MH + theoretical (u) | Cutoff distance (Å) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Thr-104r | Lys-129r | VTHIEIRD*TR | ELPLLK | DST | y3b,y4b,y7a,y6a | 9.69 | 2065.109 | 2065.129 | 16.8 |
| 2 | Lys-51a | Thr-111r | SKKTMLVPKD*ITSEATBBVAK | NTR | DST | b6a,y15a,y18,ab2b | 9.4 | 2870.379 | 2870.406 | 16.8 |
| 3 | Lys-45a | Thr-66r | KTMLVPK | TIPSHAFSD*LPNISR | DST | b5b,b4b,b11a,b4a,y4a | 7.96 | 2600.36 | 2600.339 | 16.8 |
| 4 | Lys-45a | Lys-91a | KTMLVPK | ATVmGNVRVENHTEBHBSTBYYHKS | DST | y2a,b6b,b21a,y15a | 0.72 | 3985.779 | 3985.776 | 20.4 |
| 5 | Phe-1a | Lys-101b | {FPDGEFTmQGBPEBK | BNTDYSDBIHEAIK | DST | y4a,y2a,b2b,y6a,b8b,y7a,y5a | 3.81 | 3657.408 | 3657.422 | 17.9 |
| 6 | Lys-44a | Lys-45a | SKKTMLVPK | PQK | BS3 | b3,y2,y1,y6y3,b3 | 0.24 | 1173.722 | 1173.721 | 25.4 |
| 7 | Tyr-37a | Lys-110b | AYPTPAR | IRGILESLMBD*ESSmQSLR | BS3 | b4a,y3a,b3a,b5a,y1b,b4a | 0.14 | 1284.694 | 1284.695 | 26.1 |
| 8 | Ser-19a | Ser-308r | ENKYFSK | {GQELK | BS3 | b4b,y13a,y2b,y5a | 5.97 | 3297.63 | 3297.61 | 22.7 |
| 9 | Tyr-59b | Gly-367r | DFmYK | {GQELK | BS3 | b2a,y2b,b1b | 4 | 1430.693 | 1430.687 | 40.9 |
| 10 | Tyr-26a | Lys-63a | YFSKPDAPIYQBmGBBFSR | DITSEATBBVAKAFTK | BS3 | b2,b14a,b7a,b5b,b9b,b7a | 2.82 | 4341.897 | 4341.909 | 26.1 |
| 11 | Lys-261r | Lys-287r | ELIARNTWTLKK | ADLSYPSHBBAFK | BS3 | b11a,y10b,y2a,y11a | 8.18 | 3165.612 | 3165.586 | 25.4 |
| 12 | Lys-45a | Thr-111r | SKK | VTHIEIRD*TR | EGS | y7a,y2b,b6a,y8a,b2a, | 12.69 | 1826.937 | 1826.961 | 26.5 |
| 13 | Lys-67a | Ser-298r | AFTKATVMGD*VR | NQKKIRGILESLMBD*ESSMQSLR | EGS | y7a,y3a,y4a,y6a,y9a,y8a,y5a | 4.86 | 4244.061 | 4244.082 | 26.5 |
| 14 | Lys-44a | Tyr-116r | AYPTPARSKK | NTRNLTYIDPDALK | EGS | b6b,b3a,y1a,y1b | 8.186 | 2977.502 | 2977.527 | 30.8 |
| 15 | Lys-291r | Ser-298r | NQKK | GILESLmBD*ESSmQSLRQR | BS(PEG)[5] | y17a,b8a,y1b,y16a | 2.23 | 3090.482 | 3090.475 | 32.1 |
| 16 | Lys-44a | Tyr-185r | SKKTmLVPK | LYNNGFTSVQGYAFD*GTK | BS(PEG)[5] | b8b,b2b,y7b,y1a,y1b | 4.22 | 3330.696 | 3330.681 | 36.4 |
| 17 | Lys-63a | Thr-111r | TMLVPKD*ITSEATBBVAKAFTK | NTR | BS(PEG)[5] | y1a,b9a,b10a,y21a | 0.28 | 3162.574 | 3162.573 | 32.1 |
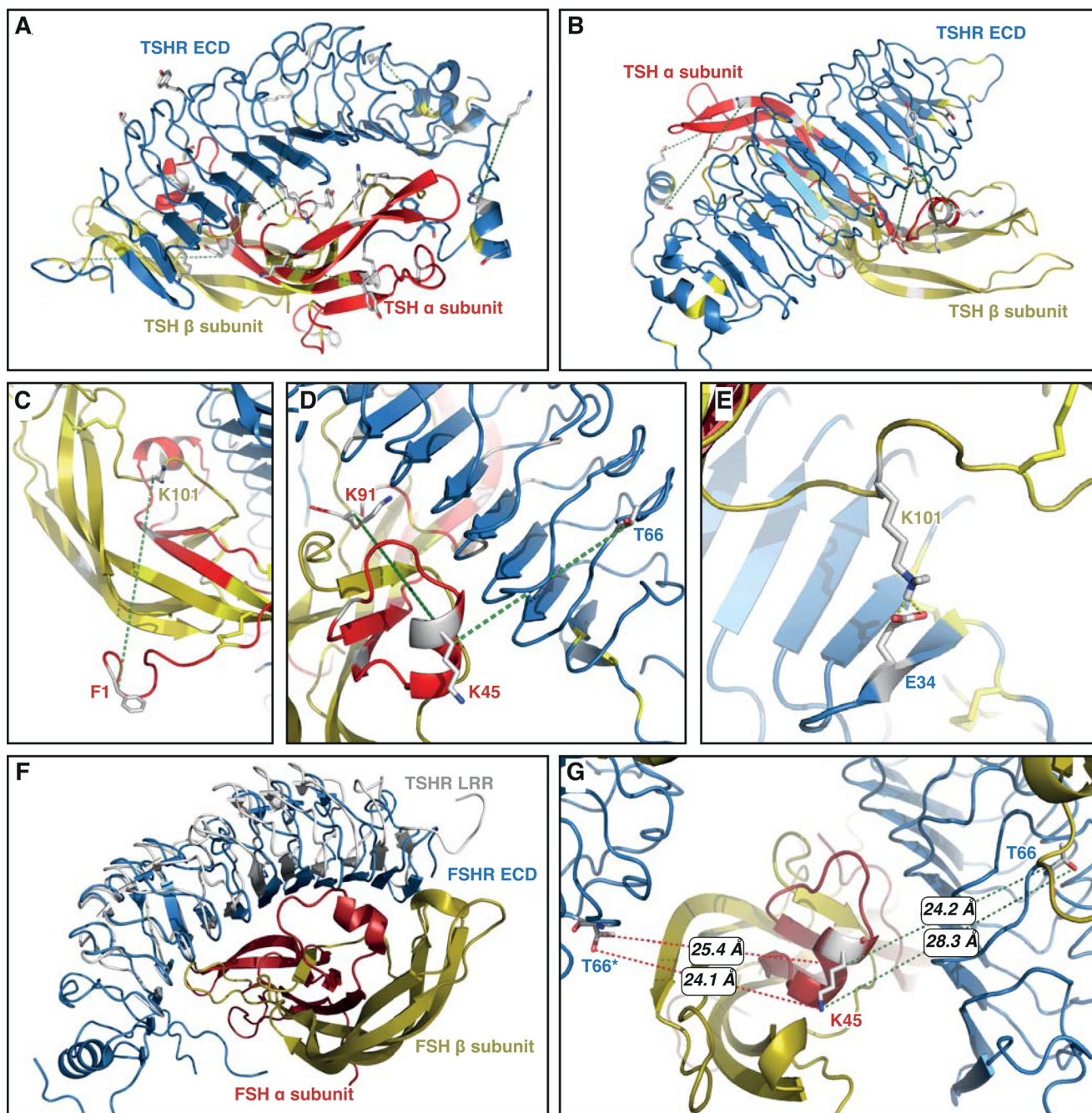
FIGURE 3. **Homology models of the TSHR-ECD (*blue*)/TSH (*red*, α-chain; *yellow*, β-chain) complex.** Cross-links (*green dotted lines*) confirm the fold (*A*) of the domains and (*B*) a similar binding mode of TSH as observed for FSH to the FSHR-ECD. *C*, the cross-link (*green line*) between the N terminus of the TSH α-chain and Lys-101 close to the C terminus of the β-chain suggests close proximity of the termini. *D*, cross-link to Lys-45 of the TSH α-chain are not met by any homology model of the TSHR-ECD·TSH complex. *E*, homology models suggest a potential TSHR-specific contact between Glu-34 of the ECD to Lys-101 of the TSH β-chain. A direct contact was observed in two homology models of the TSHR-ECD·TSH complex (one selected model depicted) supporting the feasibility of a contribution to specific TSH binding. *F*, superposition of the TSHR-LRR domain (*white*, PDB code 2xwt) at the C-terminal region with the FSHR-ECD/FSH complex (PDB code 4ay9). The reduced curvature of the TSHR-LRR resulted in an increased distance to the hormone in the N-terminal part of the TSHR template. *G*, distances between the cross-linked residues Lys-45 of the GPH α-chain and Thr-66 in the ECD to which the hormone is bound (*green lines*) and to the ECD of the adjacent ECD-hormone complex (*red lines*).

tively. A comparative model of the TSHR-ECD·TSH complex was constructed using the experimentally determined structure of the FSHR-ECD·FSH complex (PDB 4ay9; 11) as template. The N-terminal residues for this model up to Leu-57 (FSHR) were taken from the TSHR-LRR domain structure (PDB 2xwt; Ref. 25). The majority of the XL-MS restraints are consistent with this comparative model (Fig. 1*B*). Specifically, of the 1800

models in the ensemble (top 90% by score), 99% fulfill 12 or more of the 17 cross-links. Two cross-links are violated in all models (Fig. 1, IDs 3 and 4), an effect that we attribute to a conformational change induced by the cross-linker (read below). If a protein exists in multiple conformations, it is sufficient if one conformation has the amino acids in close proximity to observe the cross-link. In turn, not all conformations need

# Full-length TSHR Models

## TABLE 2
### Functional characterization of the WT and mutant TSHR

COS-7 (HEK$_{GT}$ for NFAT) cells were transiently transfected with the respective DNA constructs. Values are normalized to WT levels (basal state where applicable). Data are presented as the mean values and S.E. of at least three independent experiments, each carried out in duplicate. The pcDNA3.1/Zeo vector was used as a control. Spec. Const. Act., specific constitutive activity.

| Construct | Cell surface expression | cAMP | | Spec. Const. Act. | NFAT | | Specific binding | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | Basal | bTSH (30 mU/ml) | | Basal | bTSH (30 mU/ml) | $B_{Max}$ | $IC_{50}$ |
| | | | | | | | | $nM$ |
| WT | 100.00 | 1.00 | 13.23 ± 1.19 | 1.00 | 1.00 | 8.26 ± 1.28 | 100.0 | 5.10 ± 1.05 |
| S281C | 70.93 ± 3.11 | 2.53 ± 0.21 | 20.45 ± 1.98 | 5.08 ± 0.61 | 0.83 ± 0.07 | 3.48 ± 0.46 | 117.5 ± 4.37 | 1.45 ± 0.25 |
| S281D | 25.53 ± 2.43 | 3.97 ± 0.44 | 6.92 ± 1.31 | 23.44 ± 2.59 | 1.00 ± 0.12 | 1.14 ± 0.14 | 45.3 ± 7.61 | 0.08 ± 0.04 |
| S281I | 51.16 ± 2.16 | 11.32 ± 1.13 | 14.26 ± 1.21 | 37.35 ± 5.21 | 0.70 ± 0.05 | 0.79 ± 0.05 | 113.5 ± 1.84 | 1.08 ± 0.36 |
| I486C | 23.91 ± 1.67 | 0.86 ± 0.19 | 2.70 ± 0.34 | 4.46 ± 0.43 | 1.00 ± 0.10 | 1.37 ± 0.04 | 41.0 ± 7.03 | 0.07 ± 0.03 |
| I486D | 34.10 ± 2.87 | 3.94 ± 0.39 | 5.51 ± 0.68 | 24.16 ± 1.50 | 0.93 ± 0.07 | 0.95 ± 0.06 | 71.0 ± 6.57 | 0.19 ± 0.10 |
| I486K | 23.59 ± 2.58 | 1.91 ± 0.25 | 3.51 ± 0.61 | 5.98 ± 1.14 | 0.97 ± 0.08 | 1.07 ± 0.07 | 13.8 ± 3.61 | n.d. |
| I486S | 42.75 ± 1.90 | 6.56 ± 0.46 | 15.75 ± 2.47 | 22.37 ± 4.02 | 0.82 ± 0.07 | 1.02 ± 0.06 | 104.0 ± 2.73 | 0.86 ± 0.30 |
| T490C | 51.82 ± 3.80 | 1.03 ± 0.25 | 11.78 ± 1.24 | 1.59 ± 0.13 | 1.01 ± 0.04 | 3.07 ± 0.34 | | |
| I568T | 64.07 ± 3.69 | 13.96 ± 1.39 | 25.60 ± 2.91 | 30.19 ± 3.69 | 0.72 ± 0.05 | 2.02 ± 0.18 | 114.8 ± 1.26 | 2,23 ± 0.68 |
| S281C/I486C | 26.18 ± 1.49 | 1.83 ± 0.24 | 1.66 ± 0.24 | 5.79 ± 0.79 | 0.98 ± 0.08 | 1.03 ± 0.06 | 68.4 ± 3.55 | 0.11 ± 0.04 |
| S281C/T490C | 34.22 ± 2.13 | 1.07 ± 0.21 | 9.43 ± 1.39 | 5.57 ± 0.91 | 0.84 ± 0.11 | 1.36 ± 0.18 | | |
| S281D/I486D | 31.17 ± 2.93 | 1.10 ± 0.23 | 1.42 ± 0.27 | 8.64 ± 1.08 | 1.14 ± 0.09 | 1.42 ± 0.11 | 38.5 ± 6.67 | 0.16 ± 0.12 |
| S281D/I486K | 22.63 ± 2.63 | 1.03 ± 0.15 | 2.12 ± 0.27 | 4.27 ± 0.38 | 1.08 ± 0.09 | 1.40 ± 0.14 | 5.4 ± 1.59 | n.d. |
| S281I/I486S | 28.14 ± 1.47 | 3.14 ± 0.46 | 4.75 ± 0.48 | 19.50 ± 2.96 | 0.66 ± 0.06 | 0.67 ± 0.07 | 33.1 ± 3.38 | 0.10 ± 0.03 |
| S281I/I568T | 26.83 ± 1.66 | 12.21 ± 1.53 | 11.50 ± 1.32 | 78.35 ± 9.64 | 0.67 ± 0.12 | 1.09 ± 0.21 | 77.6 ± 1.96 | 0.11 ± 0.06 |
| pcDNA | 7.21 ± 0.82 | 0.51 ± 0.10 | 0.46 ± 0.14 | | 0.18 ± 0.03 | 0.19 ± 0.01 | | |
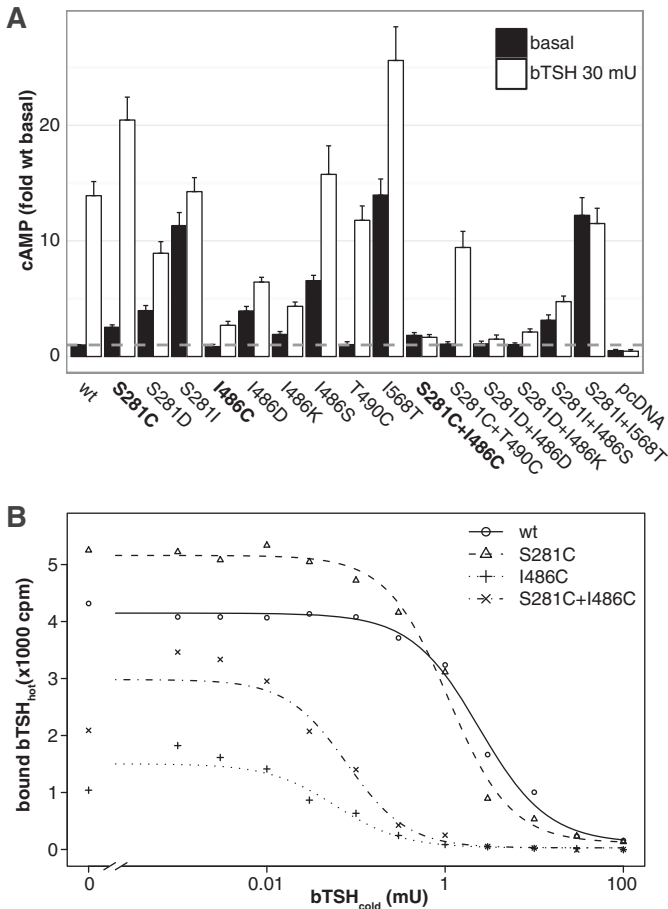


FIGURE 4. **Double mutant studies of the TSHR.** G$_s$ signaling of the TSHR and variants in the absence and presence of bTSH. The S281C/I486C and S281D/I486D double mutants do not show ligand-induced cAMP accumulation (*A*). In the case of the S281C/I486C double mutant, the transition to an activated receptor conformation is most likely hindered by a disulfide bond introduced between the two residues. The radioligand binding assay of the TSHR and variants (*B*) shows that the S281C/I486C variant is still capable of ligand binding, suggesting that the missing ligand induced receptor activation is caused by a disruption of the activation process. *mU*, milliunits.
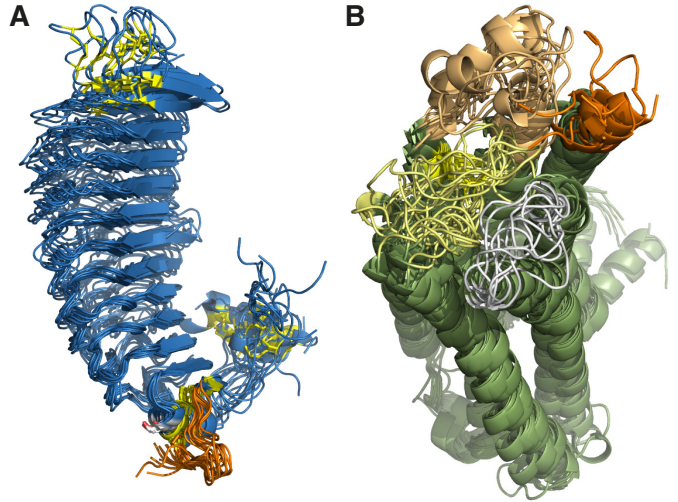


FIGURE 5. **Structural variability at the ECD/TMD interface in homology models of the TSHR.** Shown is the superposition of the best scoring homology models of the largest clusters for the TSHR-ECD (*A*) and the TSHR-TMD (*B*). The ECD models are structurally similar at the putative TMD interface located at the terminal α-helix excluding the connecting loop (depicted in *orange*), which was removed before docking. The models of the TMD, in contrast, show greater variations in the putative interface at the extracellular loops (*light orange*, ECL1; *yellow*, ECL2; *white*, ECL3).

to fulfill all cross-links simultaneously in flexible regions of the protein. Cross-links IDs 5, 8, and 9 connect flexible regions, which are assumed to be present in multiple conformations. Accordingly, we expect these to be violated in a higher fraction of the models. Cross-links within the hormone (IDs 6, 7, and 10) or the receptor (IDs 1, 11, and 15) confirm the general fold of the hormone and the ECD (Fig. 3*A*). Cross-links between the hormone and receptor close to loop 1 and 3 (α-L1/3) of the hormone's α-subunit (IDs 8 and 13) as well as to α-L2 at the opposite side of the hormone (IDs 2, 12, and 16) confirm a similar binding mode of bTSH as reported for the FSHR-ECD·FSH experimental structure (11, 37) (Fig. 3*B*).
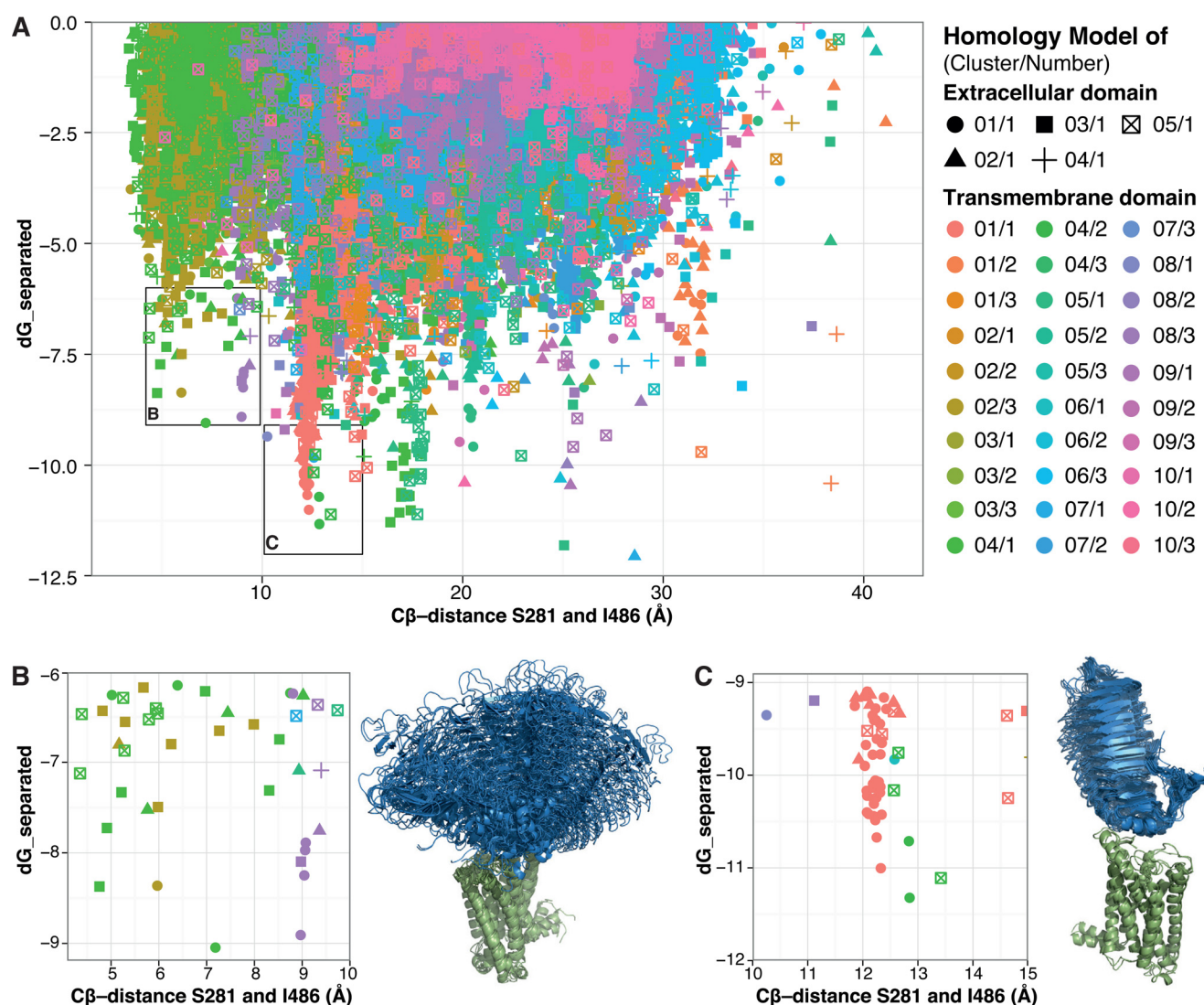
FIGURE 6. **Evaluation of docking results by interface score and Ser-281/Ile-486 Cβ distance.** Two areas (*black rectangles*) in the plot of interface score (dG separated) *versus* Cβ distance between Ser-281 and Ile-486 (*A*) were selected for reconstruction of the connecting loop. The best scoring models (dG separated $<-6$) with a Ser-281/Ile-486 Cβ distance $<10$ Å (*B*) show a broad variety of ECD/TMD orientation with a few clusters of similar orientations (up to 4 structures). Structures scoring significantly better (dG separated $<-9$) with a Ser-281/Ile-486 Cβ distance between 10 and 15 Å (*C*) display a funnel at a Ser-281/Ile-486 Cβ distance of 12.5 Å with almost all models displaying a similar ECD/TMD orientation.

### Structural Plasticity in the Curvature of the LRR Domain

Interestingly, two cross-links (Fig. 1, IDs 2 and 17) exceeded the expected maximal Cβ distance of the cross-linking reagent based on an initial model of the TSHR-ECD·TSH complex from the structure of the TSHR-LRR domain (PDB 2xwt; Ref. 25) up to Ser-234 (Ser-226 FSHR) and the HR of the FSHR-ECD·FSH structure. Superimposition of the two employed templates revealed a reduced curvature of the TSHR-LRR domain at the transition region of the templates, which results in an increased distance between the hormone and the N-terminal section of the receptor (Fig. 3F). These models also display less favorable Cβ distances for three other cross-links (IDs 3, 12, and 14). A steeper curvature was also observed in the structure of the FSHR-LRR domain (PDB 1xwd; Ref. 37). Therefore, the differences in curvature are most likely sequence specific (38) and not due the inclusion of the HR.

### Conformation of the TSHR-HR

Cross-links between the receptors HR and the TSH hormone (IDs 8, 9, and 13) confirm a significant interface between the HR with the hormone that could be important for signal transduction (Fig. 1C). It has previously been suggested that a part of the HR, including the region that is subjected to proteolytic cleavage within the TSHR, is intrinsically disordered (39). This hypothesis is supported by the FSHR-ECD crystal structure (PDB 4ay9; Ref. 11), where no density was observed for the respective region.

### A TSHR-ECD·TSH-specific Interaction between TSHR Glu-34 and TSH β-Chain Lys-101

Visual inspection of the best scoring models also suggests a potential TSHR-specific interaction at the N-terminal end of the LRR domain due to spatial proximity of the side chains of Glu-34 of the TSHR with Lys-101 of the TSH β-chain observed
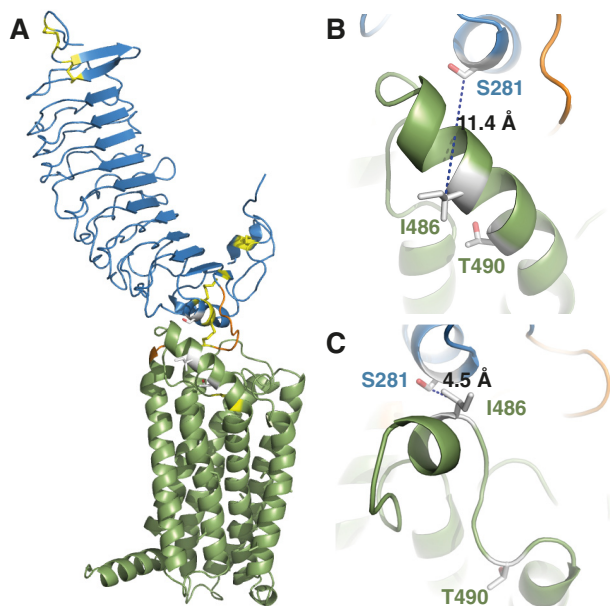
# Full-length TSHR Models



FIGURE 7. **Full-length models of the TSHR.** *A*, the best scoring full-length TSHR model of cluster one after remodeling of the connecting loop between ECD and TMD shows an almost upright orientation of the ECD toward the membrane. *B*, the conformation of ECL1 includes an extended transmembrane helix 3 in the largest cluster with Ile-486 facing away from the interface with the ECD, resulting in an increased Ser-281/Ile-486 Cβ distance. *C*, in the best model of cluster four only a small helical fragment was retained within the loop, resulting in closer proximity of Ser-281 and Ile-486. Thr-490 is located in the extended TM3 and might influence the putative transition between the extended TM3 and the loop conformation of this region during receptor activation.

in two models (Fig. 3*E*). Interestingly a TSHR mutation of Glu-34 (E34K) has been detected in patients with hypothyroidism (40). However, with no detailed binding data and only a slight impairment of $G_s$ signaling reported, the putative contribution of an Glu-34/Lys-101 interaction to binding affinity and specificity is most likely only marginal.

### Spatial Proximity between the TSH α-Chain N Terminus and TSH β-Chain Lys-101

Next to this interaction with the receptor a cross-link between Lys-101 and the N terminus of the α-chain was detected (ID 5) implying a close proximity of both termini (Fig. 3*C*). Yet the residues are within the expected distance in only 2% of the models. Because the N-terminal amino acids of the α-chain are not ordered in any of the crystallographic structures of the human GPHs (11, 37, 41–44), this region is expected to be flexible. The bovine GPH α-chain features four additional amino acids, possibly increasing flexibility in the region (45).

### Cross-links ID 3 and 4 Are Violated in All Comparative Models

DST yielded two further cross-links to the α-L2 (IDs 3 and 4) that are incompatible with all models (Fig. 3*D*). In contrast to cross-link 5, the connected residues are in structurally well defined regions. However, conformational changes in α-L2, including a disintegration of the helical fragment potentially induced by the coupling of DST, could be sufficient for the cross-link to be established. Alternatively, binding of the hormone to a second, low affinity binding site as suggested previously (13, 46, 47) could also be associated with a closer proximity of the cross-linked residues. A third explanation for the controversial cross-links is the possibility that the cross-link is not established between the hormone and the ECD it is bound to but, rather, with the ECD of the adjacent ECD-hormone complex in the putative trimer structure (48). However, analysis of this scenario reveals that the Cβ distance to the HR within the ECD-hormone complex does not differ much from the distance to the HR in the adjacent complex. Yet, side-chain orientation and surface distance are more favorable for a cross-link to the HR of the adjacent ECD·hormone complex (Fig. 3*G*).

### Identification and Verification of an ECD/ECL1 Contact between Ser-281 and Ile-486 by Double Mutant Cycle Analysis

It has been demonstrated that distant mutations result in synergistic receptor activation (49), and mutations in close proximity result in a more complex pattern dependent on the side-chain substitutions (20). Several constitutively activating mutations (CAMs) of Ser-281 (FSHR Ser-273, LHR Ser-277) in the C-terminal helix of the ECD (50–52) mark this residue as an important component of the interface with the TMD (50, 53). To confirm specific contacts of the putative ECD/TMD interface we have combined the CAM S281I with two TMD CAMs of the comparably long ECL1 (I486S, T490A) and ECL2 (I568T). For mutations of Ile-486 and Ile-568, constitutive activity has only been observed in the presence of the ECD (54). Combination of S281I (LRA 37) with I486S (LRA 22) yields a receptor with similar constitutive activity to the I486S single mutant (LRA 19) (Table 2). The absence of an additive effect suggests a shared leverage point of constitutive receptor activation and close spatial proximity. Targeted mutation of both residues to cysteines resulted in a receptor devoid of hormone-induced activation of the $G_s$ and $G_q$ signaling pathway (Fig. 4 and Table 2). Even though the ligand binding properties and cell surface expression of the double mutant are within the range of the single mutants, only the latter show ligand-induced receptor activation. Based on these observations the missing change in activity of the double mutant upon ligand binding is putatively caused by the presence of a disulfide bond between the introduced cysteines. The presence of a disulfide bond in this region critical to receptor activation most likely locks the receptor in a partially activated conformation and thus prevents signal propagation. For this bond to form, the two residues, therefore, have to be in close structural proximity in the receptor. Exchange of both residues to aspartate yields a receptor with retained ligand binding but no ligand-induced activation of G protein signaling. This observation is consistent with the notion that repulsive forces between the negatively charged aspartate side chains prevent signal propagation. Conformational changes at the ECD/ECL1 interface, including a relative repositioning of Ser-281 and Ile-486 are, therefore, a requirement for receptor activation (Table 2).

### Thr-490 and Ile-568 Are Not in Direct Contact with Ser-281

With the confirmation of the Ser-281/ECL1 contact, we tested whether Ser-281 is proximal to Thr-490 in ECL1 and Ile-568 in ECL2 (49). However, even though combining the two cysteine mutations has a detrimental effect on cell surface
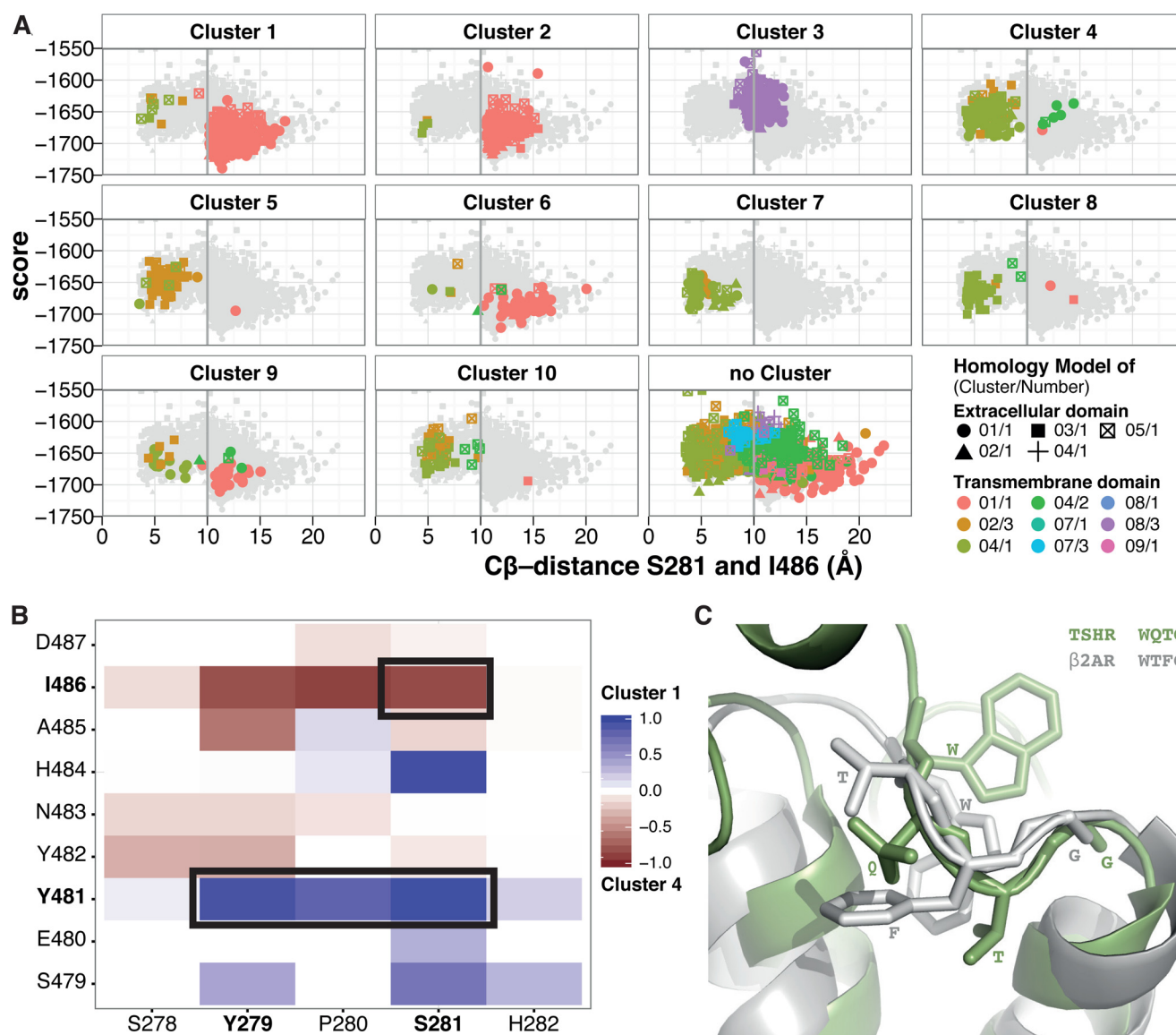
FIGURE 8. **Analysis of full-length receptor models after reconstruction of the connecting loop.** Comparison of the score *versus* Cβ distance of Ser-281 and Ile-486 after clustering (*A*) shows that the best models based on score and cluster size display a Ser-281/Ile-486 Cβ distance >10 Å. Differences in contact maps (*B*) of cluster one and four (*blue*, contact in every model of cluster one and in none of cluster four; *red*, contact only in cluster four) for the Ser-281/ECL1 interface showing that the Ser-281/Ile-486 contact is only observed in cluster four (*upper black rectangle*) and the aromatic environment of Ser-281 including Tyr-481 and Tyr-279 is only observed in cluster one (*lower black rectangle*). Superposition of the best scoring structure of cluster four (*green*) and the β2-adrenergic receptor (*white*, PDB code 2rh1) with the side chains of the W*X*FG motif depicted (*C*) shows that ECL1 of the homology model adopts a similar loop conformation in this region.

expression and ligand-induced $G_q$ signaling, no loss of $G_s$ activation was observed (Table 2). Therefore, Thr-490 is most likely not in close proximity to Ser-281. Combination of the CAMs S281I (LRA 37) and I568T (LRA 30) is synergistic yielding a receptor with increased constitutive activity (LRA 78) and high levels of basal cAMP production at 92% of the activated Wt receptor level despite a cell surface expression of only 27% compared with the WT. We conclude that S281I and I568T are unlikely to be in direct contact. Notably, combination of the CAM S281I and I568T is associated with no apparent ligand-induced receptor activation despite retained binding affinity. This suggests that the S281I/I568T double mutant adopts the conformation of the fully activated receptor lacking ligand-induced activation. A similar phenotype has been previously associated with full receptor activation (49). In that study muta-

tions in all three extracellular loops were necessary, whereas in our case substitutions at the ECD/ECL1 and ECL2/TM6 (55) interface were sufficient to enforce the activated conformation.

### Docking of ECD and TMD to Generate a Full-length Receptor Model

Whereas homology models of the ECD showed little structural variations in the presumed interface to the TMD surrounding Ser-281, models of the TMD showed greater flexibility especially in ECL1 (Fig. 5), which is, with 8–10 additional amino acids, significantly longer than in most class A GPCRs. Accordingly, the five best-scoring, representative ECD models were docked with the 30 best-scoring, representative models of the TMD. Best-scoring representative models were chosen by clustering to mimic a conformational selection process. Cluster

centers for the TMD included models derived both from TMD templates in an "active" conformation and from those in an "inactive" conformation. Analysis of the interaction energy of docked models compared with the Ser-281/Ile-486 C$\beta$ distance reveals an energy funnel at a distance of 12.5 Å (Fig. 6*A*) with a very similar orientation of the ECD toward the TMD (Fig. 6*C*) in an upright position with Ser-281 facing toward ECL1 (Fig. 7). The best scoring models with a Ser-281/Ile-486 C$\beta$ distance below 10 Å show greater diversity in the relative orientation of ECD and TMD (Fig. 6*B*). With a comparatively large Ser-281/Ile-486 C$\beta$ distance, the cluster at 17.5 Å (Fig. 6*A*) was not considered for further analysis. The addition of the hormone to the full receptor models does not result in an overlap of the hormone with the membrane in any of the structures and thus confirms the plausibility of the observed ECD orientations. To allow a free exploration of ECD/TMD orientations, the flexible linker region between ECD and TMD was constructed after completion of the docking simulation. The superior energy of the model is preserved when compared with models with a Ser-281/Ile-486 C$\beta$ distance below 10 Å (Fig. 8*A*). Strikingly, in the final model Ile-486 is part of an extended transmembrane helix 3 with the side chain facing away from the interface with the ECD (Fig. 7). In contrast the models with a shorter Ser-281/Ile-486 C$\beta$ distance lack an extended TM3, enabling a loop conformation with the Ile-486 side chain facing toward the ECD. These models also show a small helical segment within ECL1 similar to the smoothened receptor (Ref. 56; PDB 4o9r) and a conformation of the C-terminal part similar to the W*X*FG motif present in most class A GPCR structures (WQTG in all three GPHRs) to which a pivotal role in ligand-mediated receptor activation is attributed (57) (Fig. 8*B*). Mutations of Trp-488 in the TSHR result in a drastically reduced cell surface expression, suggesting a similar importance (49). The comparison of contact maps of the largest and best scoring cluster to the largest cluster with an average Ser-281/Ile-486 distance below 10 Å shows that the first cluster is consistent with placing Tyr-279 and Tyr-481 in the environment of Ser-281 as has been suggested previously (50), whereas the latter displays the experimentally determined close proximity of Ser-281 and Ile-486 (Fig. 8*C*).

### Multiple Conformations Involved in TSHR Activation

It is possible that the two observed ECL1 conformations represent different stages during GPCR activation. The extended, low energy TM3 conformation is similar to the activated state. The loop conformation observed in the fourth cluster would represent the basal state of the receptor. In this scenario Thr-490 is part of the small fragment that changes its conformation between the fold of the W*X*FG motif and an extended TM3 during activation (Figs. 7 and 8*B*). This is supported by the observation that substitution at this position to alanine, which has a higher helix probability than threonine (58), can facilitate the transition toward the activated conformation as observed in the CAM T490A. The high conservation of the region surrounding Ser-281 and ECL1 within GPHRs as well as the shared propensity for constitutive receptor activation by mutations of Ser-281 suggests an identical mechanism of activation and a shared ECD/ECL1 interface within GPHRs. The presented

modeling approach can, therefore, be easily extended to the remaining two GPHRs. The final ensemble of models offers important insights into the likely mechanism of GPHR activation. By incorporating experimental data from chemical cross-linking coupled with MS fragment analysis and targeted receptor mutation, the quality and relevance of the final model set was significantly increased and enabled the generation of the first experimentally supported full-length models of a GPHR.

*Author Contributions*—J. S., S. H., H. J., M. v. B., S. K., J. M., and R. P. conceived the study and designed the research. J. S. performed site-directed mutagenesis and functional characterization of the receptor. J. S. and M. B. M. N. expressed, purified, and cross-linked the soluble extracellular domain. M. B. M. N. and S. K. performed the mass spectrometric measurements and analysis. M. B. M. N. and V. H. performed the surface plasmon resonance experiment. J. S. and R. M. performed the molecular modeling. J. S., M. B. M. N., S. K., J. M., and R. P. wrote the paper. All authors analyzed data and approved the final version of the manuscript.

### References

1. Pierce, J. G., and Parsons, T. F. (1981) Glycoprotein hormones: structure and function. *Annu. Rev. Biochem.* **50,** 465–495
2. Braun, T., Schofield, P. R., and Sprengel, R. (1991) Amino-terminal leucine-rich repeats in gonadotropin receptors determine hormone selectivity. *EMBO J.* **10,** 1885–1890
3. Jiang, X., Dreano, M., Buckler, D. R., Cheng, S., Ythier, A., Wu, H., Hendrickson, W. A., and el Tayar, N. (1995) Structural predictions for the ligand-binding region of glycoprotein hormone receptors and the nature of hormone-receptor interactions. *Structure* **3,** 1341–1353
4. Kleinau, G., and Krause, G. (2009) Thyrotropin and homologous glycoprotein hormone receptors: structural and functional aspects of extracellular signaling mechanisms. *Endocr. Rev.* **30,** 133–151
5. Heitman, L. H., Kleinau, G., Brussee, J., Krause, G., and Ijzerman, A. P. (2012) Determination of different putative allosteric binding pockets at the lutropin receptor by using diverse drug-like low molecular weight ligands. *Mol. Cell. Endocrinol.* **351,** 326–336
6. Jäschke, H., Neumann, S., Moore, S., Thomas, C. J., Colson, A. O., Costanzi, S., Kleinau, G., Jiang, J. K., Paschke, R., Raaka, B. M., Krause, G., and Gershengorn, M. C. (2006) A low molecular weight agonist signals by binding to the transmembrane domain of thyroid-stimulating hormone receptor (TSHR) and luteinizing hormone/chorionic gonadotropin receptor (LHCGR). *J. Biol. Chem.* **281,** 9841–9844
7. Neumann, S., Pope, A., Geras-Raaka, E., Raaka, B. M., Bahn, R. S., and Gershengorn, M. C. (2012) A drug-like antagonist inhibits thyrotropin receptor-mediated stimulation of camp production in graves' orbital fibroblasts. *Thyroid* **22,** 839–843
8. Kursawe, R., and Paschke, R. (2007) Modulation of TSHR signaling by posttranslational modifications. *Trends Endocrinol. Metab.* **18,** 199–207
9. Chazenbalk, G. D., Tanaka, K., McLachlan, S. M., and Rapoport, B. (1999) On the functional importance of thyrotropin receptor intramolecular cleavage 1. *Endocrinology* **140,** 4516–4520
10. Costagliola, S., Panneels, V., and Bonomi, M. (2002) Tyrosine sulfation is required for agonist recognition by glycoprotein hormone receptors. *EMBO J.* **21,** 504–513
11. Jiang, X., Liu, H., Chen, X., Chen, P. H., Fischer, D., Sriraman, V., Yu, H. N., Arkinstall, S., and He, X. (2012) Structure of follicle-stimulating hormone in complex with the entire ectodomain of its receptor. *Proc.*

*Natl. Acad. Sci. U.S.A.* **109,** 12491–12496

12. Kalkhof, S., and Sinz, A. (2008) Chances and pitfalls of chemical cross-linking with amine-reactive *N*-hydroxysuccinimide esters. *Anal. Bioanal. Chem.* **392,** 305–312

13. Cornelis, S., Uttenweiler-Joseph, S., Panneels, V., Vassart, G., and Costagliola, S. (2001) Purification and characterization of a soluble bioactive amino-terminal extracellular domain of the human thyrotropin receptor. *Biochemistry* **40,** 9860–9869

14. Carter, P. J., Winter, G., Wilkinson, A. J., and Fersht, A. R. (1984) The use of double mutants to detect structural changes in the active site of the tyrosyl-tRNA synthetase (*Bacillus stearothermophilus*). *Cell* **38,** 835–840

15. Horovitz, A. (1996) Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1,** R121-R126

16. Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., *et al.* (2011) Rosetta3: An object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487,** 545–574

17. Schaarschmidt, J., Huth, S., Meier, R., Paschke, R., and Jaeschke, H. (2014) Influence of the hinge region and its adjacent domains on binding and signaling patterns of the thyrotropin and follitropin receptor. *PLoS ONE* **9,** e111570

18. Jaeschke, H., Schaarschmidt, J., Günther, R., and Mueller, S. (2011) The hinge region of the TSH receptor stabilizes ligand binding and determines different signaling profiles of human and bovine TSH. *Endocrinology* **152,** 3986–3996

19. Mueller, S., Kleinau, G., Jaeschke, H., Paschke, R., and Krause, G. (2008) Extended hormone binding site of the human thyroid stimulating hormone receptor: distinctive acidic residues in the hinge region are involved in bovine thyroid stimulating hormone binding and receptor activation. *J. Biol. Chem.* **283,** 18048–18055

20. Jaeschke, H., Schaarschmidt, J., Eszlinger, M., Huth, S., Puttinger, R., Rittinger, O., Meiler, J., and Paschke, R. (2014) A newly discovered TSHR variant (L665F) associated with nonautoimmune hyperthyroidism in an Austrian Family Induces constitutive TSHR activation by steric repulsion between TM1 and TM7. *J. Clin. Endocrinol. Metab.* **99,** E2051–E2059

21. Hampf, M., and Gossen, M. (2006) A protocol for combined *Photinus* and Renilla luciferase quantification compatible with protein assays. *Anal. Biochem.* **356,** 94–99

22. Müller, S. A., Pernitzsch, S. R., Haange, S.-B., Uetz, P., Bergen, von Bergen, M., Sharma, C. M., and Kalkhof, S. (2015) Stable isotope labeling by amino acids in cell culture based proteomics reveals differences in protein abundances between spiral and coccoid forms of the gastric pathogen *Helicobacter pylori*. *J. Proteomics* **126,** 34–45

23. Götze, M., Pettelkau, J., Schaks, S., Bosse, K., Ihling, C. H., Krauth, F., Fritzsche, R., Kühn, U., and Sinz, A. (2012) StavroX: a software for analyzing cross-linked products in protein interaction studies. *J. Am. Soc. Mass Spectrom* **23,** 76–87

24. Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B.-H., Das, R., Grishin, N. V., and Baker, D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **77,** 89–99

25. Sanders, P., Young, S., Sanders, J., Kabelis, K., Baker, S., Sullivan, A., Evans, M., Clark, J., Wilmot, J., Hu, X., Roberts, E., Powell, M., Núñez Miguel, R., Furmaniak, J., and Rees Smith, B. (2011) Crystal structure of the TSH receptor (TSHR) bound to a blocking-type TSHR autoantibody. *J. Mol. Endocrinol.* **46,** 81–99

26. Li, S. C., and Ng, Y. K. (2010) Calibur: a tool for clustering large numbers of protein decoys. *BMC Bioinformatics* **11,** 25

27. Song, Y., DiMaio, F., Wang, R. Y., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013) High-resolution comparative modeling with RosettaCM. *Structure* **21,** 1735–1742

28. Chaudhury, S., Berrondo, M., Weitzner, B. D., Muthu, P., Bergman, H., and Gray, J. J. (2011) Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS ONE* **6,** e22477

29. Stranges, P. B., and Kuhlman, B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22,** 74–82

30. Ho, S.-C., Goh, S.-S., Li, S., Khoo, D. H., and Paterson, M. (2008) Effects of mutations involving cysteine residues distal to the S281HCC motif at the C-terminus on the functional characteristics of a truncated ectodomain-only thyrotropin receptor anchored on glycosylphosphatidylinositol. *Thyroid* **18,** 1313–1319

31. Nagayama, Y., Nishihara, E., Namba, H., Yamashita, S., and Niwa, M. (2000) Identification of the sites of asparagine-linked glycosylation on the human thyrotropin receptor and studies on their role in receptor function and expression. *J. Pharmacol. Exp. Ther.* **295,** 404–409

32. Nicholas, H. B., Jr., Chan, S. S., and Rosenquist, G. L. (1999) Reevaluation of the determinants of tyrosine sulfation. *Endocrine* **11,** 285–292

33. Wadsworth, H. L., Chazenbalk, G. D., Nagayama, Y., Russo, D., and Rapoport, B. (1990) An insertion in the human thyrotropin receptor critical for high affinity hormone binding. *Science* **249,** 1423–1425

34. Graves, P., Pritsker, A., and Davies, T. F. (1999) Post-Translational processing of the natural human thyrotropin receptor: demonstration of more than two cleavage sites 1. *J. Clin. Endocrinol. Metab.* **84,** 2177–2181

35. Tanaka, K., Chazenbalk, G. D., McLachlan, S. M., and Rapoport, B. (2000) Evidence that cleavage of the thyrotropin receptor involves a "molecular ruler" mechanism: deletion of amino acid residues 305–320 causes a spatial shift in cleavage site 1 independent of amino acid motif 1. *Endocrinology* **141,** 3573–3577

36. Hamidi, S., Chen, C.-R., Mizutori-Sasai, Y., McLachlan, S. M., and Rapoport, B. (2011) Relationship between thyrotropin receptor hinge region proteolytic posttranslational modification and receptor physiological function. *Mol. Endocrinol.* **25,** 184–194

37. Fan, Q. R., and Hendrickson, W. A. (2005) Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **433,** 269–277

38. Jiang, X., Dias, J. A., and He, X. (2014) Structural biology of glycoprotein hormones and their receptors: insights to signaling. *Mol. Cell. Endocrinol.* **382,** 424–451

39. Mueller, S., Jaeschke, H., Günther, R., and Paschke, R. (2010) The hinge region: an important receptor component for GPHR function. *Trends Endocrinol. Metab.* **21,** 111–122

40. De Marco, G., Agretti, P., Camilot, M., Teofoli, F., Tatò, L., Vitti, P., Pinchera, A., and Tonacchera, M. (2009) Functional studies of new TSH receptor (TSHr) mutations identified in patients affected by hypothyroidism or isolated hyperthyrotrophinaemia. *Clin. Endocrinol. (Oxf.)* **70,** 335–338

41. Fox, K. M., Dias, J. A., and Van Roey, P. (2001) Three-dimensional structure of human follicle-stimulating hormone. *Mol. Endocrinol.* **15,** 378–389

42. Wu, H., Lustbader, J. W., Liu, Y., Canfield, R. E., and Hendrickson, W. A. (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein. *Structure* **2,** 545–558

43. Lapthorn, A. J., Harris, D. C., Littlejohn, A., Lustbader, J. W., Canfield, R. E., Machin, K. J., Morgan, F. J., and Isaacs, N. W. (1994) Crystal structure of human chorionic gonadotropin. *Nature* **369,** 455–461

44. Tegoni, M., Spinelli, S., Verhoeyen, M., Davis, P., and Cambillau, C. (1999) Crystal structure of a ternary complex between human chorionic gonadotropin (hCG) and two Fv fragments specific for the $\alpha$ and $\beta$-subunits. *J. Mol. Biol.* **289,** 1375–1385

45. Mueller, S., Kleinau, G., Szkudlinski, M. W., Jaeschke, H., Krause, G., and Paschke, R. (2009) The superagonistic activity of bovine thyroid-stimulating hormone (TSH) and the human TR1401 TSH analog is determined by specific amino acids in the hinge region of the human TSH receptor. *J. Biol. Chem.* **284,** 16317–16324

46. Urizar, E., Montanelli, L., Loy, T., Bonomi, M., Swillens, S., Gales, C., Bouvier, M., Smits, G., Vassart, G., and Costagliola, S. (2005) Glycoprotein hormone receptors: link between receptor homodimerization and negative cooperativity. *EMBO J.* **24,** 1954–1964

47. Zoenen, M., Urizar, E., Swillens, S., Vassart, G., and Costagliola, S. (2012) Evidence for activity-regulated hormone-binding cooperativity across glycoprotein hormone receptor homomers. *Nat. Commun.* **3,** 1007

48. Jiang, X., Fischer, D., Chen, X., McKenna, S. D., Liu, H., Sriraman, V., Yu, H. N., Goutopoulos, A., Arkinstall, S., and He, X. (2014) Evidence for follicle-stimulating hormone receptor as a functional trimer. *J. Biol. Chem.* **289,** 14273–14282

49. Kleinau, G., Jaeschke, H., Mueller, S., Raaka, B. M., Neumann, S., Paschke,

≋ASBMB

R., and Krause, G. (2008) Evidence for cooperative signal triggering at the extracellular loops of the TSH receptor. *FASEB J.* **22,** 2798 –2808

50. Jaeschke, H., Neumann, S., Kleinau, G., Mueller, S., Claus, M., Krause, G., and Paschke, R. (2006) An aromatic environment in the vicinity of serine 281 is a structural requirement for thyrotropin receptor function. *Endocrinology* **147,** 1753–1760

51. Nakabayashi, K., Kudo, M., Hsueh, A. J., and Maruo, T. (2003) Activation of the luteinizing hormone receptor in the extracellular domain. *Mol. Cell. Endocrinol.* **202,** 139 –144

52. Montanelli, L., Van Durme, J. J., Smits, G., Bonomi, M., Rodien, P., Devor, E. J., Moffat-Wilson, K., Pardo, L., Vassart, G., and Costagliola, S. (2004) Modulation of ligand selectivity associated with activation of the transmembrane region of the human follitropin receptor. *Mol. Endocrinol.* **18,** 2061–2073

53. Kleinau, G., Neumann, S., Grüters, A., Krude, H., and Biebermann, H. (2013) Novel insights on thyroid-stimulating hormone receptor signal transduction. *Endocr. Rev.* **34,** 691–724

54. Vlaeminck-Guillem, V., Ho, S.-C., Rodien, P., Vassart, G., and Costagliola, S. (2002) Activation of the cAMP pathway by the TSH receptor involves switching of the ectodomain from a tethered inverse agonist to an agonist. *Mol. Endocrinol.* **16,** 736 –746

55. Kleinau, G., Claus, M., Jaeschke, H., Mueller, S., Neumann, S., Paschke, R., and Krause, G. (2007) Contacts between extracellular loop two and transmembrane helix six determine basal activity of the thyroid-stimulating hormone receptor. *J. Biol. Chem.* **282,** 518 –525

56. Weierstall, U., James, D., Wang, C., White, T. A., Wang, D., Liu, W., Spence, J. C., Bruce Doak, R., Nelson, G., Fromme, P., Fromme, R., Grotjohann, I., Kupitz, C., Zatsepin, N. A., Liu, H., *et al.* (2014) Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* **5,** 3309

57. Klco, J. M., Nikiforovich, G. V., and Baranski, T. J. (2006) Genetic analysis of the first and third extracellular loops of the C5a receptor reveals an essential W*X*FG motif in the first loop. *J. Biol. Chem.* **281,** 12010 –12019

58. Meiler, J., Mueller, M., Zeidler, A., and Schmaeschke, F. (2001) Generation and evaluation of dimension-reduced amino acid parameter representations by artificial neural networks. *J. Mol. Model* **7,** 360 –369